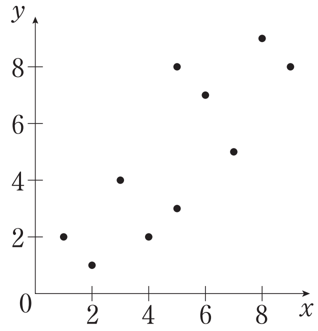
PART4 データの相関

要点① 散布図と相関関係

ある 2 つの変量の間に関係があるかを知りたいとき，その 2 つのデータの値を座標とする点をとり，視覚化する方法がある。

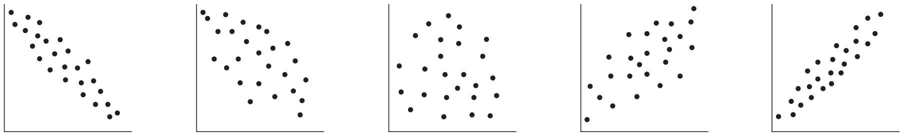
x	8	4	3	5	2	1	6	9	5	7
y	9	2	4	8	1	2	7	8	3	5



右の図は，上の表の変量 x，y の値の組 $(x,\ y)$ を座標とする点を座標平面上にとったものである。
このように，2 つの変量からなるデータを点として座標平面上に図示したものを「　　　　　」という。

2 つの変量からなるデータにおいて，一方が増えると他方も増える傾向にあるとき，2 つの変量には「　　の相関がある」という。また，一方が増えると他方が減る傾向にあるとき，2 つの変量には「　　の相関がある」という。正の相関も負の相関もみられないとき，「相関がない」という。

散布図において，点の分布が直線に近づくほど「相関が　　　」といい，広く散らばっているほど「相関が　　　」という。



負の相関（強い）　負の相関（弱い）　相関がない　正の相関（弱い）　正の相関（強い）

要点② 相関係数

ある 2 つの変量の間に関係があるかを知りたいとき，その 2 つのデータの値を用いて相関を 1 つの数値で表す方法がある。

2 つの変量 x，y からなる n 個のデータの組 $(x_1,\ y_1)$，$(x_2,\ y_2)$，……，$(x_n,\ y_n)$ について，x の偏差と y の偏差の積 $(x_k-\overline{x})(y_k-\overline{y})$ の平均値を「　　　　　」といい，「　　」と表す。

共分散

$$s_{xy}=\frac{1}{n}\{(x_1-\overline{x})(y_1-\overline{y})+(x_2-\overline{x})(y_2-\overline{y})+\cdots\cdots+(x_n-\overline{x})(y_n-\overline{y})\}$$

x と y に正の相関があるときは，共分散は正の値をとり，負の相関があるときは，共分散は負の値をとる。また，相関関係がないときは，共分散は 0 に近い値になる。

共分散は変量 x，y の単位の取り方や散らばり具合に影響を受けるため，相関の強さをみることはできない。相関関係の強弱をみるために，共分散 s_{xy} を，x，y の標準偏差 s_x，s_y の積 $s_x s_y$ で割ったものを，x，y の「　　　　　　」といい，「　」で表す。

相関係数

$$r=\frac{s_{xy}}{s_x s_y}$$

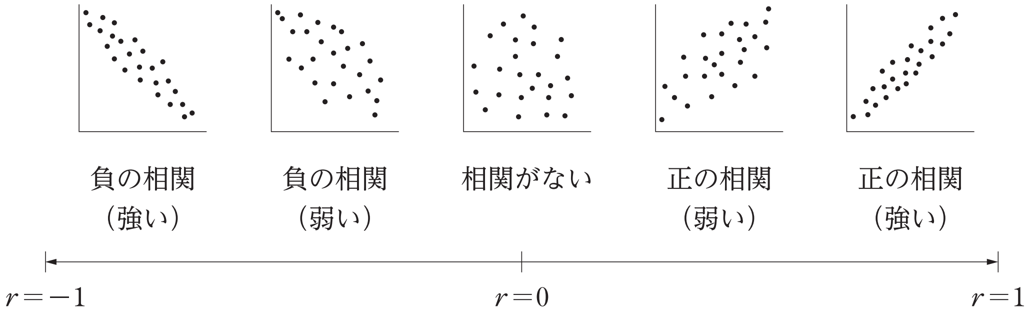
相関係数 r のとり得る値の範囲は $-1\leqq r\leqq 1$ であり

・r の値が 1 に近いほど強い正の相関がある

・r の値が -1 に近いほど強い負の相関がある

・r の値が 0 に近づくほど相関は弱い

ことが知られている。



※データの中に，他の値と極端に離れた値がある場合，相関係数の値はその影響を受けやすい。

※ 2 つの変量の間に相関関係があっても，それらの間に因果関係があるとは限らない。

例えば，ある英語と数学のテストの得点に正の相関が見られたとしても，英語の得点を高くすれば数学の得点が高くなる，または数学の得点を高くすれば英語の得点が高くなる，といった因果関係を必ずしも意味するものではない。

4 次の表は，2 つの変量 x，y からなるデータである。

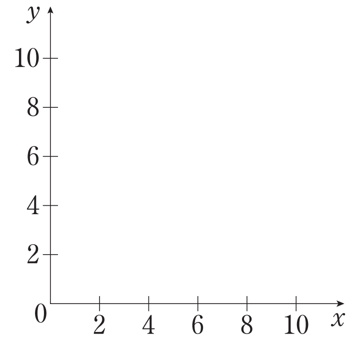
x	3	2	7	6	8	4	9	1	3	7
y	8	9	5	4	1	6	2	7	5	3

(1) 散布図をかき，x，y の間に相関があるか調べよ。また，相関がある場合は，正，負のどちらの相関があると考えられるかを答えよ。

(2) (1)の散布図から，このデータの x と y の相関係数 r に最も近いものを次の①～④から 1 つ選べ。

① -0.31　② -0.88　③ 0.25　④ 0.92

〈散布図〉



Point Pickup

散布図と相関関係

散布図における点の分布が

右上がりの直線に近い → 正の相関がある

右下がりの直線に近い → 負の相関がある

どちらの傾向もない → 相関がない

相関係数 r

相関係数 r の値が

1 に近い → 強い正の相関がある

-1 に近い → 強い負の相関がある

0 に近い → 相関がない(弱い)

第15講　PART4 確認問題

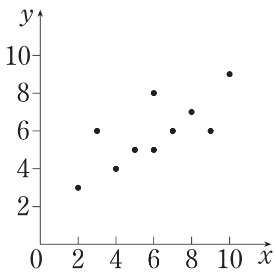
次の表は，2 つの変量 x，y からなるデータである。

x	2	3	4	5	6	6	7	8	9	10
y	3	6	4	5	5	8	6	7	6	9

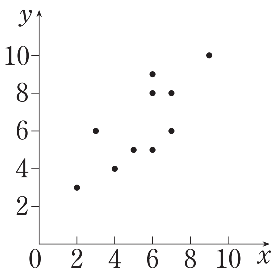
1

散布図を，次の①，②のうちから選べ。 ア

①



②



また，x と y の間にどんな関係があるといえるか。次の①～③のうちから選べ。 イ

①　正の相関があるといえる

②　負の相関があるといえる

③　相関関係はない

2

散布図から，このデータの x と y の相関係数に最も近い値を，次の①～④のうちから選べ。 ウ

①　-0.3　　　②　-0.7

③　0.3　　　④　0.7

PART4 データの相関

4 次の表は，2 つの変量 x，y からなるデータである。

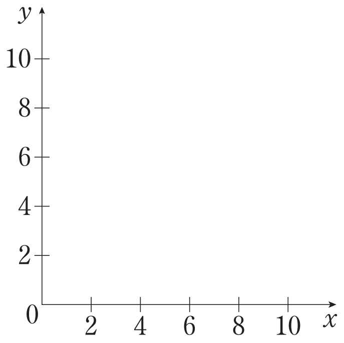
x	3	2	7	6	8	4	9	1	3	7
y	8	9	5	4	1	6	2	7	5	3

(1) 散布図をかき，x，y の間に相関があるか調べよ。また，相関がある場合は，正，負のどちらの相関があると考えられるかを答えよ。

(2) (1)の散布図から，このデータの x と y の相関係数 r に最も近いものを次の①～④から 1 つ選べ。

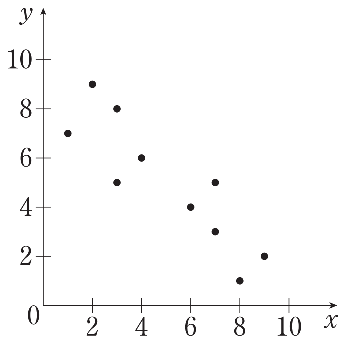
①　-0.31　　②　-0.88　　③　0.25　　④　0.92

〈散布図〉



[解答]

(1)



負の相関がある

(2) 強い負の相関がみられる。よって，相関係数 r の値は -1 に近い。

したがって　②

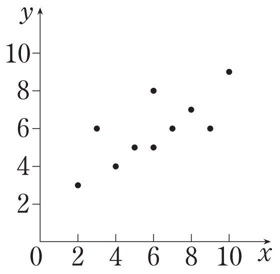
第15講 PART4 確認問題 解答

1

[正解] ア：①，イ：①

[解説]

散布図は次の図のようになる。



[A]

正しい散布図は，①

散布図から，x と y には

正の相関があるといえる。 [B]

[A]　x と y の値の組を座標 $(x,\ y)$ としてとっていく。

[B]　散布図における点の分布が
右上がりの直線に近い
→　正の相関がある
右下がりの直線に近い
→　負の相関がある

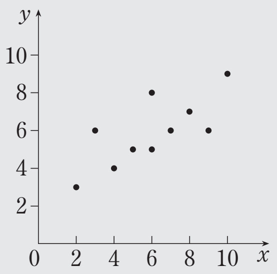
2

[正解] ウ：④

[解説]

x と y には強い正の相関が見られるので，相関係数の値は 0 よりも 1 に近い。 [C]

最も近い値は **0.7** である。

(参考)実際に相関係数を計算すると，およそ 0.74 となる。

[C]　相関係数 r について，
強い正の相関がある
→　r は 1 に近い
強い負の相関がある
→　r は -1 に近い